

At 000040

U.S.



**Europäisches
Patentamt**

**European
Patent Office**

**Office européen
des brevets**



Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

00890205.8

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 06/04/01
LA HAYE, LE

THIS PAGE BLANK (USPTO)



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 00890205.8
Demande n°:

Anmeldetag:
Date of filing: 29/06/00 ✓
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Koninklijke Philips Electronics N.V.
5621 BA Eindhoven
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:

Aufzeichnungseinrichtung zur Aufzeichnung einer Sprachinformation für eine anschließende Offline-
Spracherkennung

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:
State:
Pays:

Tag:
Date:
Date:

Aktenzeichen:
File no.
Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:

/

Am Anmeldetag benannte Vertragsstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du dépôt:

Bemerkungen:
Remarks:
Remarques:

THIS PAGE BLANK (USP 10)

PHAT000040 EP-P

- 1 -

Aufzeichnungseinrichtung zur Aufzeichnung einer Sprachinformation
für eine anschließende Offline-Spracherkennung

- 5 Die Erfindung bezieht sich auf eine Aufzeichnungseinrichtung zum Aufzeichnen einer Sprachinformation eines Diktats und zum anschließenden Abgeben der aufgezeichneten Sprachinformation des Diktats an eine Spracherkennungseinrichtung zur Offline-Spracherkennung mit Empfangsmitteln zum Empfangen der Sprachinformation des Diktats und mit Aufzeichnungsmitteln zum Aufzeichnen der empfangenen Sprachinformation des
- 10 Diktats in einer Aufzeichnungsbetriebsart der Aufzeichnungseinrichtung und mit Abgabemitteln zum Abgeben der aufgezeichneten Sprachinformation des Diktats in einer Abgabebetriebsart der Aufzeichnungseinrichtung an die Spracherkennungseinrichtung, die zum Erkennen einer der abgegebenen Sprachinformation zuzuordnenden Textinformation ausgebildet ist, wobei die Qualität der erkannten Textinformation von der Qualität der
- 15 empfangenen Sprachinformation abhängig ist.

- Ein solche Aufzeichnungseinrichtung gemäß der vorstehend im ersten Absatz angegebenen Gattung ist aus dem Dokument GB-A-2,323,693 bekannt und durch einen
- 20 Computerserver eines Offline-Spracherkennungsservices gebildet. Ein Benutzer kann zur Verwendung des Offline-Spracherkennungsservices den Computerserver mit einem Telefon anrufen und ein Diktat in das Telefon sprechen. Empfangsmittel des Computerservers sind mit der Telefonleitung verbunden und durch eine Harddisc gebildete Aufzeichnungsmittel speichern das empfangene Sprachsignal.
- 25 Nach Beendigung des Diktats durch den Benutzer wird das aufgezeichnete Sprachsignal zur Offline-Spracherkennung an eine Spracherkennungseinrichtung abgegeben. Die Spracherkennungseinrichtung erkennt die dem Sprachsignal zuzuordnende
- Textinformation, worauf ein Mitarbeiter des Offline-Spracherkennungsservices die offensichtlichen Fehler in der erkannten Textinformation korrigiert. Anschließend wird die
- 30 erkannte und korrigierte Textinformation per Email an den Benutzer des Offline-Spracherkennungsservices übermittelt und die in Anspruch genommene Leistung dem Benutzer verrechnet.

PHAT000040 EP-P

- 2 -

Bei der bekannten Aufzeichnungseinrichtung hat sich als Nachteil erwiesen, dass der Benutzer des Offline-Spracherkennungsservices ein gegebenenfalls langes Diktat diktiert und keinerlei Feedback erhält, ob die Qualität des empfangenen und aufgezeichneten Sprachsignals ausreichend gut ist, um das anschließend abgegebene Sprachsignal mit der

5 Spracherkennungseinrichtung erfolgreich verarbeiten zu können. Wenn es bei einem Benutzer des Offline-Spracherkennungsservices einmal passiert, dass das aufgezeichnete Sprachsignal seines Diktats – beispielsweise auf Grund einer schlechten Telefonleitung – nicht von der Spracherkennungseinrichtung ausgewertet werden kann, dann wird dieser Benutzer das Offline-Spracherkennungsservice nicht mehr verwenden, was

10 nachteiligerweise einen großen finanziellen Schaden für den Betreiber des Offline-Spracherkennungsservices bedeutet.

Als weiterer Nachteil der bekannten Aufzeichnungseinrichtung hat sich ergeben, dass bei einer schlechten Qualität der erkannten Textinformation der Mitarbeiter des Offline-Spracherkennungsservices eine Vielzahl von offensichtlichen Fehlern ausbessern muss.

15 Somit können die Kosten für das Offline-Spracherkennungsservice wesentlich reduziert werden, wenn der Benutzer des Offline-Spracherkennungsservice bereits während des Diktats auf die schlechte Qualität der empfangenen Sprachinformation hingewiesen wird.

20 Die Erfindung hat sich zur Aufgabe gestellt, eine Aufzeichnungseinrichtung zu schaffen, bei der der Benutzer eines Offline-Spracherkennungsservices bereits während des Diktats Feedback erhält, ob die Qualität des empfangenen Sprachsignals ausreicht, um mit der nachgeschalteten Spracherkennungseinrichtung bei einer Offline-Spracherkennung eine Textinformation mit einer ausreichenden hohen Qualität zu erkennen.

25 Diese Aufgabenstellung wird bei einer Aufzeichnungseinrichtung gemäß der in dem ersten Absatz angegebenen Gattung durch das Vorsehen von Sprachqualitätsprüfmitteln, die zum Prüfen ausgebildet sind, ob die Qualität der in der Aufzeichnungsbetriebsart empfangenen Sprachinformation ausreicht, um bei einer Verarbeitung der in der Abgabebetriebsart mit den Abgabemitteln abgegebenen Sprachinformation mit der

30 Spracherkennungseinrichtung eine vorgegebene Qualität der erkannten Textinformation zu erhalten, und durch das Vorsehen von Feedbackmitteln gelöst, die zum Abgeben einer Feedbackinformation in der Aufzeichnungsbetriebsart der Aufzeichnungseinrichtung

PHAT000040 EP-P

- 3 -

ausgebildet sind, welche Feedbackinformation das Ergebnis der Prüfung der Sprachqualitätsprüfmittel wiedergibt.

Hierdurch ist der Vorteil erhalten, dass der Benutzer der Aufzeichnungseinrichtung bereits während des Diktats Feedback erhält, ob die Qualität der empfangenen

- 5 Sprachinformation zur erfolgreichen Verarbeitung mit einer Spracherkennungseinrichtung ausreichen wird.

- Es kann erwähnt werden, dass ein Diktiergerät bekannt ist, das eine Aussteuerungsanzeige aufweist, anhand der der Benutzer die Aussteuerung des aufgezeichneten Sprachsignals erkennt. Der Benutzer kann hierauf zum Erhalt einer
- 10 besseren Aussteuerung des aufgezeichneten Sprachsignals lauter oder leiser sprechen. Dieses Überwachen nur eines von vielen Parametern des empfangenen Sprachsignals genügt jedoch keinesfalls, um die vorstehend angegebene Aufgabe zuverlässig zu lösen.

- Weiters kann erwähnt werden, dass eine Spracherkennungseinrichtung zur Abarbeitung eines Online-Spracherkennungsverfahrens bekannt ist, bei der ein Benutzer ein Diktat in
- 15 ein Mikrofon diktieren kann und nach einer kurzen Verarbeitungszeit mit einem Monitor den erkannten Text angezeigt erhält. Zusätzlich wird mit dem Monitor die Aussteuerung des verarbeiteten Sprachsignals dargestellt. Wenn der Benutzer durch viele Fehler in dem erkannten Text erkennt, dass die Qualität des erkannten Texts schlecht ist, dann kann der Benutzer die Darstellung der Aussteuerung des verarbeiteten Sprachsignals dazu nutzen,
- 20 seine Lautstärke beim Diktieren entsprechend zu verändern oder die Audioanpassung des Mikrofons der Spracherkennungseinrichtung neu durchzuführen.

- Bei dem bekannten Diktiergerät und der bekannten Spracherkennungseinrichtung zur Online-Spracherkennung wird nur ein Parameter des empfangenen und verarbeiteten Sprachsignals dargestellt, was in keiner Weise einer Prüfung der Qualität der empfangenen
- 25 Sprachinformation entspricht, ob diese für einen anschließenden Spracherkennungsvorgang geeignet ist. Zusätzlich erfolgt bei dem bekannten Diktiergerät und der bekannten Spracherkennungseinrichtung über die reine Darstellung der Aussteuerung des verarbeiteten Sprachsignals hinaus kein weiterer Hinweis, ab welchem Aussteuerungswert die Qualität der empfangenen Sprachinformation nicht ausreichend gut für eine
- 30 anschließende Offline-Spracherkennung ist, was bei einer reinen Aufzeichnung eines Diktats zur späteren Audiowiedergabe oder bei einer Online-Spracherkennung auch nicht nötig ist. Somit war durch das bekannte Diktiergerät und die bekannte

PHAT000040 EP-P

- 4 -

Spracherkennungseinrichtung zur Online-Spracherkennung weder die Aufgabenstellung für die erfindungsgemäße Aufzeichnungseinrichtung noch die erfindungsgemäße Lösung dieser Aufgabenstellung bekannt.

5 Gemäss den Maßnahmen des Anspruchs 2 und des Anspruchs 3 ist der Vorteil erhalten, dass die Qualität des von der Aufzeichnungseinrichtung aufzuzeichnenden Sprachsignals bereits bei der Aufzeichnung bezüglich eines ausreichend guten Signal- zu Rauschverhältnisses und eines ausreichend guten Pegels geprüft wird.

10 Gemäss den Maßnahmen des Anspruchs 4 ist der Vorteil erhalten, dass geprüft wird, wie schnell der Benutzer die Worte des Diktats spricht, und dass der Benutzer bei einer zu hohen Sprechgeschwindigkeit durch eine Feedbackinformation aufgefordert wird langsamer zu sprechen, um eine entsprechend hohe Qualität der erkannten Textinformation zu erhalten.

15 Gemäss den Maßnahmen des Anspruchs 5 ist der Vorteil erhalten, dass geprüft wird, wie deutlich bzw. verständlich der Benutzer die Worte des Diktats spricht, und dass der Benutzer bei einer zu undeutlichen Aussprache der Wörter durch eine Feedbackinformation aufgefordert wird deutlicher bzw. verständlicher zu sprechen, um eine entsprechend hohe Qualität der erkannten Textinformation zu erhalten.

20 Gemäss den Maßnahmen des Anspruchs 6 ist der Vorteil erhalten, dass eine Aufzeichnungseinrichtung dazu ausgebildet ist, dem Benutzer während des Diktats einen Hinweis auf den Grund der niedrigen Qualität der empfangenen Sprachinformation zu geben, um durch darauffolgende Maßnahmen durch den Benutzer eine Aufzeichnung der Sprachinformation in einer für eine nachfolgende Spracherkennung ausreichend hohen Qualität zu erreichen.

25

Die Erfindung wird im Folgenden anhand von zwei Ausführungsbeispielen näher erläutert, auf die die Erfindung aber nicht beschränkt ist.

Figur 1 zeigt einen Anrufbeantworter als Aufzeichnungseinrichtung, von dem eine aufgezeichnete Sprachinformation an eine Spracherkennungseinrichtung abgebar ist.

30 Figur 2 zeigt ein tragbares Diktiergerät, das an eine Spracherkennungseinrichtung anschliessbar ist und das Sprachqualitätsprüfmittel und drei Leuchtdioden aufweist, die gegebenenfalls einen Hinweis auf den Grund der niedrigen Qualität der empfangenen

PHAT000040 EP-P

- 5 -

Sprachinformation geben.

Die Figur 1 zeigt ein Offline-Spracherkennungssystem 1 zum Abarbeiten eines Offline-
5 Spracherkennungsverfahrens, das ein Telefon 2, einen Anrufbeantworter 3, einen
Spracherkennungs-Computer 4 und einen Benutzer-Computer 5 aufweist. Das Telefon 2
und der Benutzer-Computer 5 stehen dem Benutzer des Offline-Spracherkennungssystems
1 zur Verfügung und sind über ein Telefon/Datennetz NET mit dem Anrufbeantworter 3
und dem Spracherkennungs-Computer 4 verbunden.

10 Der Spracherkennungs-Computer 4 arbeitet eine Spracherkennungssoftware - wie
beispielsweise die Spracherkennungssoftware SpeechMagic der Fa.Philips - ab und bildet
hierbei eine Spracherkennungseinrichtung für ein Offline-Spracherkennungsverfahren. Die
Spracherkennungseinrichtung ist zum Empfangen von digitalen Sprachdaten SD(SI) einer
abgegebenen Sprachinformation SI und zum Erkennen einer der empfangenen
15 Sprachinformation SI zuzuordnenden Textinformation TI ausgebildet, wie dies seit langem
bekannt ist.

Die Qualität der erkannten Textinformation TI kann durch einen Erkenn-Qualitätswert
beschrieben werden, der als Verhältnis der in der erkannten Textinformation TI enthaltenen
richtig erkannten Wörtern zu allen in der erkannten Textinformation TI enthaltenen
20 Wörtern ermittelt werden kann. Die Qualität der von der Spracherkennungseinrichtung
erkannten Textinformation TI ist von vielen Einflussfaktoren anhängig. Solche
Einflussfaktoren sind beispielsweise der Trainingszustand der
Spracherkennungseinrichtung und die Anzahl der von der Spracherkennungseinrichtung
erkennbaren und in einem Basislexikon enthaltenen Wörter. Einer dieser Einflussfaktoren
25 ist aber auch die Qualität der von der Spracherkennungseinrichtung empfangenen
Sprachinformation SI, da selbst die beste Spracherkennungseinrichtung zu einer qualitativ
schlechten Sprachinformation SI keine Textinformation TI mit einem guten Erkenn-
Qualitätswert erkennen kann.

Die Qualität der Sprachinformation SI kann bereits beim Diktat durch zu starke
30 Hintergrundgeräusche oder durch ein schlechtes Mikrofon des Telefons 2 oder aber auch
erst bei der Übertragung der Sprachinformation SI durch schlechte
Übertragungseigenschaften der Telefonleitung des Telefon/Datennetzes NET nachteilig

PHAT000040 EP-P

- 6 -

beeinflusst werden. Die Qualität eines die Sprachinformation SI enthaltenden analogen elektrischen Sprachsignals SS(SI) kann hierbei durch das Signal zu Rauschverhältnis des Sprachsignals SS(SI) und durch den Pegel des Sprachsignals SS(SI) beschrieben werden.

Wenn die in dem Sprachsignal SS(SI) enthaltene Sprachinformation SI von der
5 Spracherkennungseinrichtung zur Erkennung der Textinformation TI ausgewertet werden soll, dann kann die Qualität der Sprachinformation SI für das Offline-Spracherkennungsverfahren auch durch weitere Kennwerte beschrieben werden. Einer dieser Kennwerte ist ein Sprechgeschwindigkeitskennwert, der die Geschwindigkeit der Abfolge der in der Sprachinformation SI enthaltenen Wörter angibt, worauf nachfolgend
10 noch näher eingegangen ist.

Sowohl der Spracherkennungs-Computer 4 als auch der Benutzer-Computer 5 arbeiten eine E-Mail-Software – wie beispielsweise Lotus Notes der Fa.IBM – ab und sind mit dem Telefon/Datennetz NET verbunden. Der Spracherkennungs-Computer 4 sendet als Abschluss des Offline-Spracherkennungsverfahrens die erkannte Textinformation TI als
15 Attachement eines E-Mails an den Benutzer-Computer 5 des Benutzers des Offline-Spracherkennungssystems 1.

Der Anrufbeantworter 3 bildet eine Aufzeichnungseinrichtung zum Aufzeichnen der an den Anrufbeantworter 3 abgegebenen Sprachinformation SI eines Diktats und zum anschließenden Abgeben der aufgezeichneten Sprachinformation SI des Diktats. Hierfür
20 weist der Anrufbeantworter 3 Empfangsmittel 6, Aufzeichnungsmittel 7 und Abgabemittel 8 auf.

Die Empfangsmittel 6 enthalten einen Telefonstecker 9 und eine Telefoninterfacestufe 10. Der Anrufbeantworter 3 ist über den Telefonstecker 9 mit dem Telefon/Datennetz NET verbunden. Der Benutzer des Offline-Spracherkennungssystems 1 kann mit seinem Telefon
25 2 die Telefonnummer des Anrufbeantworters 3 wählen und die Sprachinformation SI seines Diktats, das er gerne als erkannte Textinformation TI mit seinem Benutzer-Computer 5 empfangen würde, in das Telefon 2 sprechen. Hierbei ist die
—Telefoninterfacestufe 10 zum Empfangen und Abgeben der dem Telefonstandard entsprechenden Signale und Spannungen ausgebildet, wie dies seit langem bekannt ist. Das
30 Sprachsignal SS(SI) der Sprachinformation SI wird somit von dem Telefon 2 über das Telefon/Datennetz NET, den Telefonstecker 9 und die Telefoninterfacestufe 10 an die Aufzeichnungsmittel 7 abgegeben.

PHAT000040 EP-P

- 7 -

Die Aufzeichnungsmittel 7 enthalten einen Analog-Digital-Wandler, um das analoge Sprachsignal SS(SI) der Sprachinformation SI in digitale Sprachdaten SD(SI) zu wandeln. Die Aufzeichnungsmittel 7 enthalten weiters eine Harddisc, mit der die Sprachdaten SD(SI) in einer Aufzeichnungsbetriebsart des Anrufbeantworters 3 aufgezeichnet werden.

- 5 Die Aufzeichnungsmittel 7 bilden einen Teil der Abgabemittel 8 und sind in einer Abgabebetriebsart des Anrufbeantworters 3 zum Wiedergeben bzw. Abgeben der aufgezeichneten Sprachdaten SD(SI) ausgebildet. Die Abgabemittel 8 des Anrufbeantworters 3 enthalten weiters eine Datenverarbeitungsstufe 11, mit der die wiedergegebenen Sprachdaten SD(SI) gemäß einem Codierverfahren codiert werden, das
- 10 dem Spracherkennungs-Computer 4 die weitere Verarbeitung der Sprachdaten SD(SI) erleichtert. Über einen Datenanschluss 12 der Abgabemittel 8 werden die verarbeiteten Sprachdaten SD(SI) von dem Anrufbeantworter 3 an den Spracherkennungs-Computer 4 abgegeben.

- Der Anrufbeantworter 3 weist nunmehr Sprachqualitätsprüfmittel 13 auf, die zum
- 15 Prüfen ausgebildet sind, ob die Qualität der in der Aufzeichnungsbetriebsart empfangenen Sprachinformation SI ausreicht, um bei einer Verarbeitung der in der Abgabebetriebsart von den Abgabemitteln 8 abgegebenen Sprachinformation SI mit der Spracherkennungseinrichtung eine vorgegebene Qualität der erkannten Textinformation TI zu erreichen. Hierfür sind die Sprachqualitätsprüfmittel 13 zum Prüfen des Signal zu
- 20 Rauschverhältnisses des Sprachsignals SS(SI), des Pegels des Sprachsignals SS(SI) und des Sprechgeschwindigkeitskennwertes der Sprachinformation SI ausgebildet.

- Zum Prüfen des Signal zu Rauschverhältnisses wird von Zeitabschnitten des Sprachsignals SS(SI) mit jeweils 50ms der Energieinhalt der Spektralanteile ermittelt und eine Häufigkeitsverteilung der ermittelten Energieinhalte erstellt. Umso breiter die Kurve
- 25 der auf diese Weise ermittelten Häufigkeitsverteilung ist, desto geringer ist das Signal zu Rauschverhältnis des Sprachsignals SS(SI). Dem Fachmann sind weitere vergleichbare Verfahren zur Ermittlung des Signal zu Rauschverhältnisses eines elektrischen Signals bekannt.

- Bei einem zu geringen Signal zu Rauschverhältnis sind relativ große Rauschanteile in
- 30 dem Sprachsignal SS(SI) enthalten und führen bei der Analog-Digital-Wandlung des Sprachsignals SS(SI) zu einer Störung, weshalb die in den Sprachdaten SD(SI) enthaltene Sprachinformation SI nicht mehr mit der von dem Benutzer gesprochenen

PHAT000040 EP-P

- 8 -

Sprachinformation SI übereinstimmt. Die in einem solchen qualitativ schlechten Sprachsignal SS(SI) mit einem zu geringen Signal zu Rauschverhältnis enthaltene Sprachinformation SI kann von der Spracherkennungseinrichtung nur mit einem relativ schlechten Erkenn-Qualitätswert erkannt werden.

- 5 Zum Prüfen des Pegels des Sprachsignals SS(SI) werden die ermittelten Energieinhalte der Spektralanteile addiert, wobei Spikes der Energieinhalte unberücksichtigt bleiben. Bei einem zu geringen Pegel des Sprachsignals SS(SI) wirken sich dem Sprachsignal SS(SI) überlagerte Störsignale wesentlich stärker aus, als wenn das Sprachsignal SS(SI) einen ausreichend hohen Pegel aufweist. Ein zu geringer Pegel des Sprachsignals SS(SI) kennzeichnet daher eine schlechte Qualität des Sprachsignals SS(SI).

- 10 Zum Prüfen der Sprechgeschwindigkeit ermitteln die Sprachqualitätsprüfmittel 13 den Sprechgeschwindigkeitskennwert der Sprachinformation SI. Hierfür ermitteln in den Sprachqualitätsprüfmitteln 13 vorgesehene Zeitmessmittel die Zeitabstände benachbarter Abschnitte des Sprachsignals SS(SI), die hohe Energieinhalte aufweisen. Je geringer die auf diese Weise ermittelten Zeitabstände der Abschnitte des Sprachsignals SS(SI) sind, desto rascher folgen Vokale in der Sprachinformation SI aufeinander und umso rascher hat der Benutzer des Offline-Spracherkennungssystems 1 gesprochen.

- 15 Die Qualität der Sprachinformation SI ist für die Verarbeitung entsprechend dem Offline-Spracherkennungsverfahren schlecht, wenn der ermittelte Sprechgeschwindigkeitskennwert kennzeichnet, dass der Benutzer relativ schnell gesprochen hat. Eine solche Sprachinformation SI ist von der Spracherkennungseinrichtung nur schlecht erkennbar, weshalb die von der Spracherkennungseinrichtung in diesem Fall erkannte Textinformation TI eine schlechte Qualität und einen nur geringen Erkenn-Qualitätswert aufweist.

- 25 Das Prüfen des Sprachsignals SS(SI) durch die Sprachqualitätsprüfmittel 13 erfolgt während einer Zeitspanne von jeweils 10 Sekunden, worauf als Ergebnis der Prüfung der Qualität der empfangenen Sprachinformation SI eine Sprachqualitätsinformation QI an Feedbackmittel 14 des-Anrufbeantworters 3 abgegeben wird. Durch das Prüfen des Sprachsignals SS(SI) während der Zeitspanne ist erreicht, dass eine nur kurzfristig schlechte Qualität des Sprachsignals SS(SI) durch beispielsweise ein Knacksen in der Telefonleitung zu keinem niedrigen Wert der Sprachqualitätsinformation QI führt.

Die Feedbackmittel 14 sind bei in dem Anrufbeantworter 3 aktivierter

PHAT000040 EP-P

- 9 -

Aufzeichnungsbetriebsart und beim Empfang eines niedrigen Wertes der Sprachqualitätsinformation QI zum Abgeben einer Feedbackinformation FI an den Benutzer des Offline-Spracherkennungssystems 1 ausgebildet. Die Feedbackinformation FI kennzeichnet hierbei das Ergebnis der Prüfung der Sprachqualitätsprüfmittel 13.

- 5 Hierdurch ist der Vorteil erhalten, dass der Benutzer des Offline-Spracherkennungssystems 1 bereits während des Diktierens eine Feedbackinformation FI erhält, wenn die Qualität der in dem Anrufbeantworter 3 empfangenen Sprachinformation SI zu schlecht ist, um bei dem darauffolgenden Offline-Spracherkennungsverfahren eine erkannte Textinformation TI mit einer ausreichend hohen Qualität zu ermitteln. Der
- 10 Benutzer kann daher vorteilhafterweise unmittelbar auf die Feedbackinformation FI reagieren und es ist vermieden, dass das gesamte Diktat des Benutzers wegen einer zu niedrigen Qualität der Sprachinformation SI nicht verwendet werden kann.

- Die Funktionsweise des Offline-Spracherkennungssystems 1 und die Vorteile des Anrufbeantworters 3 sind im Folgenden anhand eines Anwendungsbeispiels näher
- 15 erläutert. Es sein angenommen, dass der Benutzer des Offline-Spracherkennungssystems 1 ein Arzt ist, der einen Befund diktieren möchte. Weiters sei angenommen, dass der Arzt bereist einmal eine Identifizierungsroutine des Offline-Spracherkennungssystems 1 abgearbeitet hat und dass - auf in der Figur 1 nicht näher dargestellte Weise - die Telefonnummer des Telefons 2 des Arztes gemeinsam mit der Kreditkartennummer des
- 20 Arztes, zur Verrechnung der Offline-Spracherkennungsserviceleistung, in dem Offline-Spracherkennungssystem 1 gespeichert wurde.

- Der Arzt wählt nunmehr die Telefonnummer des Anrufbeantworters 3 und hört den Anrufbeantwortertext ab, der die Benutzung des Online-Spracherkennungssystems 1 kurz erklärt. Durch die in dem Telefonprotokoll übertragene Telefonnummer des Telefons 2
- 25 identifizieren die Telefoninterfacemittel 10 den Anrufer, wodurch die Verrechnung der Offline-Spracherkennungsserviceleistung ermöglicht ist. Anschließend wird die Aufzeichnungsbetriebsart des Anrufbeantworters 3 aktiviert und der Arzt beginnt mit dem Sprechen beziehungsweise Diktieren des Befundes. Hierauf wird die Sprachinformation SI
- der gesprochenen Wörter als Sprachsignal SS(SI) mit den Empfangsmitteln 6 empfangen
- 30 und als Sprachdaten SD(SI) mit der Harddisc der Aufzeichnungsmittel 7 gespeichert.

Während der Arzt diktiert, kommt seine Assistentin in das Zimmer und öffnet das Fenster, worauf der Straßenlärm als relativ starkes Hintergrundgeräusch in der von dem

PHAT000040 EP-P

- 10 -

Anrufbeantworter empfangenen Sprachinformation SI enthalten ist. Der auf sein Diktat konzentrierte Arzt nimmt hiervon keine Notiz und diktiert seinen Befund weiter.

Die Sprachqualitätsprüfmittel 13 stellen hierauf als Ergebnis der laufend durchgeführten Prüfung des Signal zu Rauschverhältnisses, des Pegels und der Sprechgeschwindigkeit fest, dass das Signal zu Rauschverhältnis des Sprachsignals SS(SI) wesentlich geringer geworden ist und geben eine entsprechende Sprachqualitätsinformation QI an die Feedbackmittel 14 ab. Die Feedbackmittel 14 geben hierauf einen in den Feedbackmitteln 14 aufgezeichneten Text als Feedbackinformation FI wieder, mit dem der Benutzer auf zu laute Hintergrundgeräusche in seinem Raum hingewiesen wird. Ein dem wiedergegebenen Text entsprechendes Sprachsignal SS(FI) wird über den Telefonstecker 9 und das Telefon/Datennetz NET an das Telefon 2 abgegeben. Mit dem Feedbacktext FI wird der Benutzer weiters gebeten, die Hintergrundgeräusche zu reduzieren, da sonst die erkannte Textinformation TI viele Fehler aufweisen wird.

Hierdurch ist der Vorteil erhalten, dass der Arzt sofort auf das Problem des zu lauten Hintergrundgeräusches hingewiesen wurde und eine Vielzahl von Fehlern in der erkannten Textinformation TI des ab dem Öffnen des Fensters diktieren Diktats vermieden werden konnten. Durch den Feedbacktext FI aufmerksam gemacht, schließt der Arzt das Fenster und setzt das Diktat fort, wodurch das Problem mit dem zu lauten Hintergrundgeräusch gelöst ist.

Nachdem der Arzt das Diktat fertig diktiert hat beendet er das Telefonat, worauf der Anrufbeantworter von der Aufzeichnungsbetriebsart in die Abgabebetriebsart schaltet und die aufgezeichneten Sprachdaten SD(SI) des Diktats des Arztes an den Spracherkennungs-Computer 4 abgibt. Wenn während der Abgabe der Sprachdaten SD(SI) ein weiterer Benutzer des Offline-Spracherkennungssystems 1 den Anrufbeantworter 3 anruft, dann wird die Abgabebetriebsart unterbrochen und zu einem späteren Zeitpunkt fortgesetzt.

Die durch den Spracherkennungs-Computer 4 gebildete Spracherkennungseinrichtung erkennt hierauf die den abgegebenen Sprachdaten SD(SI) zuzuordnende Textinformation TI und gibt diese als ein Attachment eines E-Mails an den Benutzer-Computer 5 des Arztes ab. Der Arzt erhält somit den zu seinem Diktat erkannten Text per E-Mail und die Kosten dieses Offline-Spracherkennungsservices werden von seinem Kreditkartenkonto abgebucht.

Gemäß einem weiteren Anwendungsbeispiel des Offline-Spracherkennungssystems 1

PHAT000040 EP-P

- 11 -

wählt der Arzt die Telefonnummer des Anrufbeantworters 3, wobei in diesem Fall die Übertragungsqualität der Telefonleitung des Telefon/Datennetzes NET relativ schlecht ist. Dies wird von den Sprachqualitätsprüfmitteln 14 bereits bei der Prüfung des Sprachsignals SS(SI) der ersten Wörter des Diktats des Arztes durch ein geringes Signal zu

- 5 Rauschverhältnis und einen geringen Pegel des Sprachsignals SS(SI) festgestellt.

Vorteilhafterweise erhält der Arzt daher von den Feedbackmitteln 14 bereits nach den ersten Wörtern des Diktats eine Feedbackinformation FI, mit der er gebeten wird, das Telefonat zu beenden und neuerlich anzurufen, da die Qualität der Telefonleitung sehr schlecht ist. Hierdurch ist neuerlich vermieden, dass der Arzt das gesamte Diktat in das

10 Telefon 2 diktiert und auf Grund der schlechten Qualität des erkannten Textes mit dem Offline-Spracherkennungssystem 1 unzufrieden ist.

Durch die unmittelbare Feedbackinformation FI der Aufzeichnungseinrichtung kann vorteilhafterweise zusätzlich vermieden werden, dass ein Mitarbeiter des Offline-Spracherkennungssystems 1, der den von dem Spracherkennungs-Computer 4 erkannten

15 Text auf offensichtliche Fehler überprüft bevor dieser per E-Mail an den Benutzer-Computer 5 übermittelt wird, sehr viele Fehler korrigieren musste. Es konnte daher durch das unmittelbare Feedback an den Benutzer zusätzlich Arbeitszeit des Mitarbeiters gespart werden, womit das Offline-Spracherkennungsservice kostengünstiger durchgeführt werden kann.

20 In der Figur 2 ist ein in der Hand des Benutzers haltbares digitales Diktiergerät 15 dargestellt, das ebenfalls eine Aufzeichnungseinrichtung für eine Offline-Spracherkennung bildet. Die von dem Benutzer gesprochene Sprachinformation SI des Diktats wird von Empfangsmitteln 16 empfangen, die ein Mikrofon 17 und eine Signalverarbeitungsstufe 18 enthalten.

25 In einer Aufzeichnungsbetriebsart des Diktiergeräts 15 wird das von der Signalverarbeitungsstufe 18 abgegebene Sprachsignal SS(SI) der Sprachinformation SI mit Aufzeichnungsmitteln 19 gespeichert, die einen Analog-Digital-Wandler und ein Flash-RAM mit einer Stützbatterie als Speichermedium enthalten.

Die Aufzeichnungsmittel 19 bilden einen Teil von Abgabemitteln 20 und geben in einer

30 Abgabebetriebsart des Diktiergeräts 15 die die aufgezeichnete Sprachinformation SI enthaltenden digitalen Sprachdaten SD(SI) an eine Datenverarbeitungsstufe 21 der Abgabemittel 20 ab. Die Datenverarbeitungsstufe 21 ist zum Verarbeiten der

PHAT000040 EP-P

- 12 -

wiedergegebenen Sprachdaten SD(SI) und zum Abgeben der verarbeiteten Sprachdaten SD(SI) an einen Datenanschluss 22 ausgebildet.

Das Diktiergerät 15 ist mit seinem Datenanschluss 22 an einen Spracherkennungs-Computer 23 anschließbar, der eine Spracherkennungssoftware abarbeitet. Die von dem
5 Spracherkennungs-Computer 23 zu der von dem Diktiergerät 15 abgegebenen Sprachinformation SI erkannte Textinformation TI wird von dem Spracherkennungs-Computer 4 an einen Monitor 24 abgegeben und mit diesem dargestellt.

Das Diktiergerät 15 weist nunmehr weiters Sprachqualitätsprüfmittel 13 auf, deren Funktionsweise mit der Funktionsweise der Sprachqualitätsprüfmittel 13 gemäß Figur 1
10 übereinstimmt. Die von den Sprachqualitätsprüfmitteln 13 ermittelte Sprachqualitätsinformation QI ist an Feedbackmittel 25 und die Aufzeichnungsmittel 19 abgebar.

Die Feedbackmittel 25 geben eine Einschaltspannung als erste Feedbackinformation FI1 an eine Leuchtdiode 26 ab, wenn die Sprachqualitätsprüfmittel 13 ein zu geringes Signal-
15 zu Rauschverhältnis des Sprachsignals SS(SI) festgestellt haben. Die Feedbackmittel 25 geben weiters die Einschaltspannung als zweite Feedbackinformation FI2 an eine Leuchtdiode 27 ab, wenn die Sprachqualitätsprüfmittel 13 einen zu geringen Pegel des Sprachsignals SS(SI) festgestellt haben. Die Feedbackmittel 25 geben weiters die Einschaltspannung als dritte Feedbackinformation FI3 an eine Leuchtdiode 28 ab, wenn die
20 Sprachqualitätsprüfmittel 13 eine zu hohe Sprechgeschwindigkeit des Benutzers anhand des Sprachsignals SS(SI) festgestellt haben.

Im Folgenden ist die Funktionsweise des Diktiergeräts 15 anhand eines Anwendungsbeispiels näher erläutert. Gemäß dem Anwendungsbeispiel ist angenommen, dass sich der Benutzer des Diktiergeräts 15 auf Reisen befindet und ein Diktat mit dem
25 Diktiergerät 15 aufzeichnen möchte, um das aufgezeichnete Diktat nach seiner Ankunft zu Hause mit seinem Spracherkennungs-Computer 23 verarbeiten zu lassen und mit dem Monitor 24 die erkannte Textinformation TI dargestellt zu erhalten.

Hierfür aktiviert der Benutzer die Aufzeichnungsbetriebsart des Diktiergeräts-15-und- beginnt mit der Aufzeichnung des Diktats. Da der Benutzer sehr in Eile ist, diktiert er sehr
30 schnell, weshalb die Sprechgeschwindigkeit der zu diesem Zeitpunkt von dem Diktiergerät 15 aufgezeichnete Sprachinformation SI zu hoch ist, um bei dem Offline-Spracherkennungsverfahren mit dem Spracherkennungs-Computer 23 einen ausreichend

PHAT000040 EP-P

- 13 -

hohen Erkenn-Qualitätswert der erkannten Textinformation TI zu erhalten.

Die Sprachqualitätsprüfmittel 13 geben hierauf eine entsprechende Sprachqualitätsinformation QI an die Feedbackmittel 25 ab, welche ihrerseits die dritte Feedbackinformation FI3 an die Leuchtdiode 28 abgeben. Somit beginnt die Leuchtdiode 28 zu leuchten und weist den Benutzer darauf hin, dass er zu schnell spricht und die aufgezeichneten Sprachdaten SD(SI) nicht zur Offline-Spracherkennung geeignet sein werden.

Vorteilhafterweis erkennt der Benutzer durch das Leuchten der Leuchtdiode 28, dass er die Wörter des Diktats zu schnell spricht und spricht hierauf langsamer. Somit erlischt die Leuchtdiode 28 und die ab diesem Zeitpunkt aufgezeichneten Sprachdaten SD(SI) sind für die anschließende erfolgreiche Verarbeitung mit dem Spracherkennungs-Computer 23 geeignet.

Die Aufzeichnungsmittel 19 zeichnen die von den Sprachqualitätsprüfmitteln 13 ermittelte Sprachqualitätsinformation QI den jeweiligen Sprachdaten SD(SI) zugehörig ab, um diese in der Abgabebetriebsart des Diktiergeräts 15 gemeinsam mit den Sprachdaten SD(SI) an den Spracherkennungs-Computer 23 abzugeben.

Hierdurch ist der Vorteil erhalten, dass der Spracherkennungs-Computer 23 bei der Verarbeitung von Sprachdaten SD(SI), die eine Sprachinformation SI mit einer schlechte Qualität aufweisen, zum Ergreifen von geeigneten Maßnahmen ausgebildet sein kann, um diese Sprachdaten SD(SI) mit einem möglichst guten Ergebnis der Erkenn-Qualität zu verarbeiten. Solche Maßnahmen können beispielsweise durch eine geeignete Vor-Verarbeitung der empfangenen digitalen Sprachdaten SD(SI) gebildet sein.

Es kann erwähnt werden, dass als Offline-Spracherkennungsverfahren in diesem Zusammenhang ein Spracherkennungsverfahren verstanden wird, bei dem der Benutzer nicht unmittelbar beim Diktieren die der Sprachinformation SI zugeordnete erkannte Textinformation TI angezeigt erhält, wie dies bei Online-Spracherkennungsverfahren der Fall ist. Ein solches Online-Spracherkennungsverfahren ist beispielsweise mit der Spracherkennungssoftware FreeSpeech2000 der Fa.Philips realisiert.

Es kann erwähnt werden, dass es durch den Aufzeichnungs- und Abgabevorgang der Sprachinformation SI mit den Aufzeichnungsmitteln zu keiner zusätzlichen Verschlechterung der Qualität der empfangenen Sprachinformation SI kommt, da die Sprachinformation SI als digitale Sprachdaten SD(SI) aufgezeichnet werden. Daher kann

PHAT000040 EP-P

- 14 -

tatsächlich – wie vorstehend beschrieben - durch das Prüfen der Qualität des empfangenen analogen Sprachsignals SS(SI) mit den Sprachqualitätsprüfmitteln auf die Qualität der letztendlich von der Spracherkennungseinrichtung empfangenen und verarbeiteten digitalen Sprachdaten SD(SI) geschlossen werden.

- 5 Es kann erwähnt werden, dass die Qualität der Sprachinformation SI für das Offline-Spracherkennungsverfahren auch durch einen weiteren Kennwerte, nämlich einen Verständlichkeits-Kennwert, beschrieben werden kann. Durch eine statistische Analyse der zeitlichen Abfolge von Frequenzanteilen des Sprachsignals SS(SI) mit den Sprachqualitätsprüfmitteln kann festgestellt werden, ob in dem Sprachsignal SS(SI) eine
- 10 Sprachinformation SI oder lediglich Geräusche oder Musik enthalten sind. Der Verständlichkeits-Kennwert kann auch durch eine alternative Analyse des Sprachsignals SS(SI) ermittelt werden, bei der unterschiedliche Merkmalsgrößen (mittlere Energie, Spektrum, Nulldurchgänge, Pitch,...) des Sprachsignals SS(SI) verkoppelt werden, um das Sprachsignal SS(SI) zu klassifizieren.
- 15 Wenn ein Schwellwert für den ermittelten Verständlichkeits-Kennwert zur Unterscheidung, ob in dem Sprachsignal SS(SI) eine Sprachinformation SI enthalten ist oder nicht, relativ hoch festgelegt wird, dann wird auch bereits eine von dem Benutzer undeutlich bzw. schlecht verständlich ausgesprochene Sprachinformation SI als Geräusch klassifiziert. Auf diese Weise ist es ermöglicht, dass eine erfindungsgemäße
- 20 Aufzeichnungseinrichtung zum Abgeben einer Feedbackinformation FI an den Benutzer ausgebildet ist, die den Benutzer auf seine undeutliche Artikulation bei der Aussprache des Diktats hinweist. Hierauf wird der Benutzer deutlicher bzw. verständlicher sprechen, wodurch sich vorteilhafterweise die Qualität der erkannten Textinformation TI verbessern wird.

PHAT000040 EP-P

- 15 -

Patentansprüche:

1. Aufzeichnungseinrichtung zum Aufzeichnen einer Sprachinformation eines Diktats und zum anschließenden Abgeben der aufgezeichneten Sprachinformation des Diktats an eine Spracherkennungseinrichtung zur Offline-Spracherkennung mit
- 5 Empfangsmitteln zum Empfangen der Sprachinformation des Diktats und mit Aufzeichnungsmitteln zum Aufzeichnen der empfangenen Sprachinformation des Diktats in einer Aufzeichnungsbetriebsart der Aufzeichnungseinrichtung und mit Abgabemitteln zum Abgeben der aufgezeichneten Sprachinformation des Diktats in einer Abgabebetriebsart der Aufzeichnungseinrichtung an die Spracherkennungseinrichtung, die
- 10 zum Erkennen einer der abgegebenen Sprachinformation zuzuordnenden Textinformation ausgebildet ist, wobei die Qualität der erkannten Textinformation von der Qualität der empfangenen Sprachinformation abhängig ist, und mit Sprachqualitätsprüfmitteln zum Prüfen, ob die Qualität der in der Aufzeichnungsbetriebsart empfangenen Sprachinformation ausreicht, um bei einer Verarbeitung der in der
- 15 Abgabebetriebsart mit den Abgabemitteln abgegebenen Sprachinformation mit der Spracherkennungseinrichtung eine vorgegebene Qualität der erkannten Textinformation zu erhalten, und mit Feedbackmitteln zum Abgeben einer Feedbackinformation in der Aufzeichnungsbetriebsart, die das Ergebnis der Prüfung der Sprachqualitätsprüfmittel
- 20 wiedergibt.
2. Aufzeichnungseinrichtung gemäß Anspruch 1, wobei die Empfangsmittel zum Empfangen eines die Sprachinformation enthaltenden Sprachsignals ausgebildet sind und wobei die Sprachqualitätsprüfmittel zum Prüfen des Signal zu Rauschverhältnisses des empfangenen Sprachsignals ausgebildet sind und wobei bei einem zu geringen Signal zu
- 25 Rauschverhältnis von den Feedbackmitteln eine entsprechende Feedbackinformation abgebar ist.
3. Aufzeichnungseinrichtung gemäß Anspruch 1, wobei die Empfangsmittel zum Empfangen eines die Sprachinformation enthaltenden Sprachsignals ausgebildet sind und wobei die Sprachqualitätsprüfmittel zum Prüfen des Pegels des empfangenen
- 30 Sprachsignals ausgebildet sind und wobei bei einem zu geringen Pegel von den Feedbackmitteln eine entsprechende Feedbackinformation abgebar ist.
4. Aufzeichnungseinrichtung gemäß Anspruch 1, wobei die Sprachqualitätsprüfmittel

PHAT000040 EP-P

- 16 -

beim Prüfen der empfangenen Sprachinformation zum Ermitteln der Sprechgeschwindigkeit des Benutzers ausgebildet sind und wobei bei einer zu hohen Sprechgeschwindigkeit von den Feedbackmitteln eine entsprechende Feedbackinformation abgebar ist.

- 5 5. Aufzeichnungseinrichtung gemäß Anspruch 1, wobei die Sprachqualitätsprüfmittel beim Prüfen der empfangenen Sprachinformation zum Ermitteln der Verständlichkeit bzw. Deutlichkeit der von dem Benutzer gesprochenen Wörter des Diktats ausgebildet sind und wobei bei einer nicht verständlichen bzw. zu undeutlichen Aussprache der Wörter durch den Benutzer von den Feedbackmitteln eine entsprechende Feedbackinformation abgebar
- 10 ist.

6. Aufzeichnungseinrichtung gemäß Anspruch 1, wobei die Feedbackmittel zum Abgeben einer Feedbackinformation ausgebildet sind, die dem Benutzer einen Hinweis gibt, wie die Qualität der empfangenen Sprachinformation durch Maßnahmen des Benutzers verbessert werden kann.

- 15 7. Aufzeichnungseinrichtung gemäß Anspruch 1, wobei die Aufzeichnungseinrichtung durch ein in der Hand des Benutzers tragbares Diktiergerät gebildet ist.

8. Aufzeichnungseinrichtung gemäß Anspruch 1, wobei die Empfangsmittel zum Empfangen der Sprachinformation an eine Telephonleitung beziehungsweise Datenleitung anschließbar sind.

PHAT000040 EP-P

- 17 -

Zusammenfassung:

Aufzeichnungseinrichtung zur Aufzeichnung einer Sprachinformation
für eine anschließende Offline-Spracherkennung

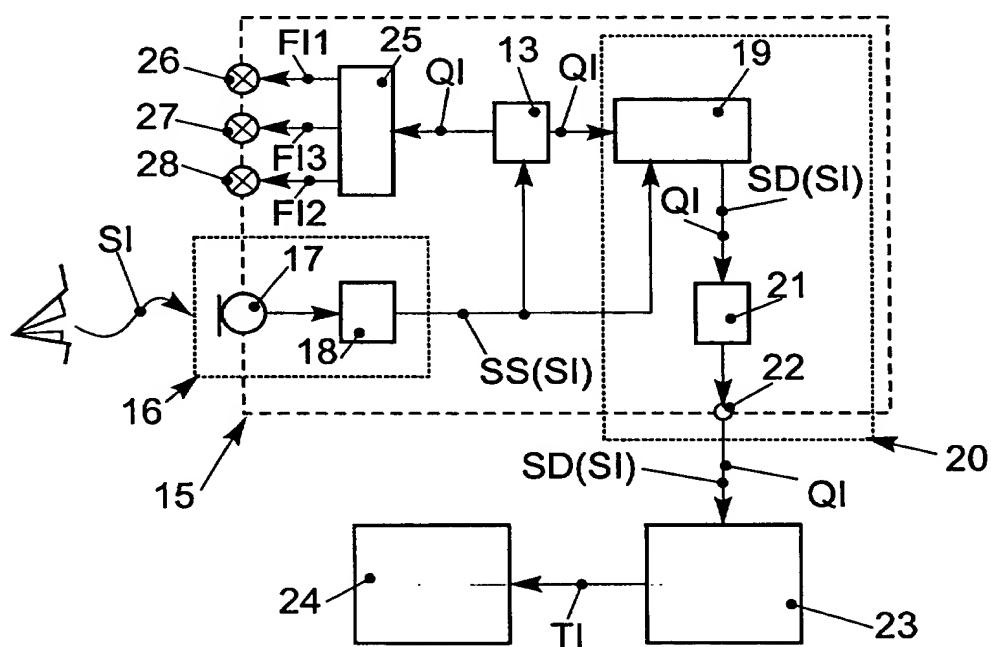
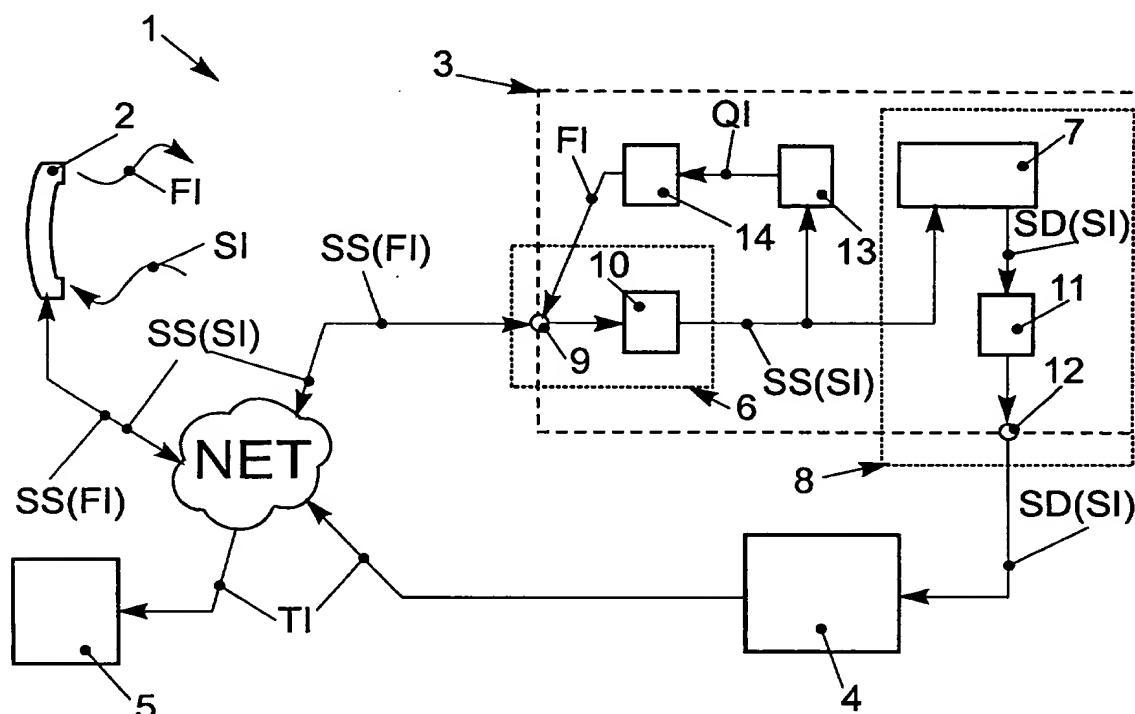
5

- Bei einer Aufzeichnungseinrichtung (3; 15) zum Aufzeichnen einer Sprachinformation (SI) eines Diktats und zum anschließenden Abgeben der aufgezeichneten Sprachinformation (SI) des Diktats an eine Spracherkennungseinrichtung (4; 23) zur Offline-Spracherkennung sind Empfangsmitteln (6; 16) zum Empfangen der Sprachinformation (SI) des Diktats und sind Aufzeichnungsmitteln (7; 19) zum Aufzeichnen der empfangenen Sprachinformation (SI) des Diktats in einer Aufzeichnungsbetriebsart der Aufzeichnungseinrichtung (3; 15) und sind Abgabemitteln (8; 20) zum Abgeben der aufgezeichneten Sprachinformation (SI) des Diktats in einer Abgabebetriebsart der Aufzeichnungseinrichtung (3; 15) an die Spracherkennungseinrichtung (4; 23) vorgesehen, die zum Erkennen einer der abgegebenen Sprachinformation (SI) zuzuordnenden Textinformation (TI) ausgebildet ist, wobei die Qualität der erkannten Textinformation (TI) von der Qualität der empfangenen Sprachinformation (SI) abhängig ist, wobei nunmehr Sprachqualitätsprüfmitteln (13) zum Prüfen, ob die Qualität der in der Aufzeichnungsbetriebsart empfangenen Sprachinformation (SI) ausreicht, um bei einer Verarbeitung der in der Abgabebetriebsart mit den Abgabemitteln (8; 20) abgegebenen Sprachinformation (SI) mit der Spracherkennungseinrichtung (4; 23) eine vorgegebene Qualität der erkannten Textinformation (TI) zu erhalten, und Feedbackmitteln (14; 25, 26, 27, 28) zum Abgeben einer Feedbackinformation (FI; FI1, FI2, FI3) in der Aufzeichnungsbetriebsart vorgesehen sind, welche Feedbackinformation (FI; FI1, FI2, FI3) das Ergebnis der Prüfung der Sprachqualitätsprüfmittel (13) wiedergibt.

(Figur 1)

THIS PAGE BLANK (USPTO)

1/1



THIS PAGE BLANK (USPTO)